

Wed May 7 15:00:47 2003

pct-us03-05226-16.rge

Page 2

Intimate interaction with human epithelial cells
Mol. Microbiol. 14 (2), 217-233 (1994)
JOURNAL
MEDLINE
PUBMED
7830568
2 (bases 1 to 4317)
REFERENCE
AUTHORS
JOURNAL
St. Gene, J.W.
Submitted (17-JUN-1994) Joseph W. St. Gene, Molecular Microbiology,
Washington University School of Medicine, 660 South Euclid Ave.,
St. Louis, MO 63110, USA
COMMENT
On Jul 12, 1996 this sequence version replaced gi:562013.
FEATURES
Location/Qualifiers
1..4317
/organism="Haemophilus influenzae"
/strain="N187"
/db_xref="taxon:727"
61..4242
/gene="hap"
61..4242
/gene="hap"
/standard_name="Haemophilus adhesion and penetration"
/note="hap"
/codon_start=1
/transl_table=11
/product="adhesion and penetration protein"
/protein_id="AAB03707.1"
/db_xref="gi:1408569"
/translation="MKTVPRNLFACISLIGVSAWAGHTFYGIDQYRDAENK
GFTYGAONIKYKNGGOLVGTSMKTFKPMDFSVNSALAYEMOYVSAHVAGY
TPVDGASGNPDORHRTYKVRKMYKKNLHEDDYNPELHEVTEAALDMS
KNNGSTISRTKYPKRVIGSGQKQFNRDQKQVAGAHILFAGTNQAGNGY
STLGDVVRAGYGLPLTASGSGSSPMPTDYAEKQWILNLEGPFGKENG
OLVRSYFDELFERDLRTSLYTRAGNAYVYISGNDGSGITQKSGISEIKITAM
SLPLEKRVNPRYDGNITSPRLNNGELTENDQKGLIPASDINAGAGLYEFG
NFTVSPNMOVWAGISVENSSTWVKGVEYEDRSKIGKSTLVQAKGKNKIS
VGDGKILLEQADQGNKQAFSEIGVSGKQVLANQDLPEDKRYFEGFAGDLN
GSLFKRIQNDGAMVNHNTQANVTITGESVLENGNINKLRYREILAVGR
WFEEDKNNHNRNLKYPTEDETRLLSGGKTLKQIDQTKGKFEFGPRPHYN
HLNKRSEMEGIPQSEIYWDHMTNTEFKENPDIGKGSVAVRNSSTEGWNTSYN
ANATGYVNOONTICTRSDMTGLTQCKXDIDTQKYNSEIPQIOLNGSINLQDA
NYKGLAKINGNTLTHNSQPLSNMNAQIGIRLSDNSTATYONANLNVHITDSQ
FESKNSHSHQOGDGTIVLENAITWPEBDITLQNLINSTITLNSASISSNT
PERRLEITPTLSAEHRNLTIVNGKUSGGGTQFTSSLEHGRKSKLISDAEDY
ILSVRTGEPEPTLEQLIVESKDNQPSDLEKLTLENDVAGALRYKLVDEGFR
LHNPILKEDELNDIVADQAEKRLTLEKQVEPTATQGEKRVSRRAAAPPDILPD
OSLNLALKEKQELTAEQKSKATKRVSRKAYFSDPLIDQSLFALFALFVIAAPQ
OSEKRLIAEAEKQKQKDLISRSNALSLSATVNSMISVODEDLFTVDOASA
VMTNIDKRRYDAPFAYOQKNTLRQIGVOKLNGRIGAVSHSRSPNTDEQV
NHATLIMSGFAOYOMDLOFGVANGTGISAKSANGRIKRAKINGVANSYOFK
LQGLIQPEYGNARYFIERNYQSEEVKVPISLAENRYNAGIRVYTTEDPNTSVK
PYRFVYVNSNANQOTVNLVLAQPPGRYQWQEVKLKELHPLQISAFISQSGSO
LGRQNVGKLYRW"

BASE COUNT 1496 a 776 c 890 g 1155 t
ORIGIN

Query Match 62.0%; Score 2995; DB 1; Length 4317;
Best Local Similarity 82.4%; Pred. No. 0;
Matches 3582; Conservative 0; Mismatches 645; Indels 120; Gaps 8;

QY 253 TCATATAGCGTTTACACAGTATTTTAAATGCAAAATTCCTAATTAATAAAT 312
Db 1 TCATATAGCGTTTACACAGTATTTTAAATGCAAAATTCCTAATTAATAAAT 60
QY 313 ATGAAAAAACTGATTTTCTGTAATTTTAAACGCTGATTCATAGGAGTACTA 372
Db 61 ATGAAAAAACTGATTTTCTGTAATTTTAAACGCTGATTCATAGGAGTACTA 120
QY 373 TCGCAGGCTGGAGGCTGACACTATTTTGGAGTACATCAATATATATGATATTT 432
Db 121 TCGCAGGCTGGAGGCTGACACTATTTTGGAGTACATCAATATATGATATTT 180

Db 181 GCCGAGATATTAAGGAGTTCACACTTGGGCTCAAAATATTAAGTTTAAACAAACA 240
QY 493 GGAATTTAGTTGGACATCAATGACAAAGCCCAATGATTTTCCGTGCTGCG 552
Db 241 GGGCAATTTAGTTGGACATCAATGACAAAGCCCAATGATTTTCCGTGCTGCG 300
QY 553 CGAAATGGGCTGGGCTGATTTGGGCGCATCATATATTTGAGTGGACATTAATGA 612
Db 301 CGTACAGGCTGGGCTGATTTGGGCGCATCATATATTTGAGTGGGCGACATTAATGA 360
QY 613 GCGTATACCAATTTGATTTTGGTCTGTAAGGACAAATCTGTATCAACTGTTTACT 672
Db 361 GGAATATACCAATTTGATTTTGGTCTGTAAGGACAAATCTGTATCAACTGTTTACT 420
QY 673 TATATAATTTGTGAAGGATATTTATTAATAACGATCAACGATCCTTATGAGAAAGAC 732
Db 421 TATATAATTTGTGAAGGATATTTATTAATAACGATCAACGATCCTTATGAGAAAGAC 480
QY 733 TACCAACACGACGCTTACATATTTTGTTCGAGGACACCCCAATCGATATGACTCT 792
Db 481 TACCAATATCCAGTATACATATTTTGTTCGAGGACGCTCCCAATGATATGACTCTG 540
QY 793 GATATGAAGGCAACAAATATACAGATAGAGAGAAATATCCGAAACGCTGCTATCGGC 852
Db 541 AATATGAATGCGAGTACTTATTCAGATGACAAATATCCAGAAAGTGTTCGATCGGC 600
QY 853 TCCGGGTGGCAGTTTGGGGAACGATCAAAACAGGCGACCAAGTTGCGGCGCATAT 912
Db 601 TCTGAGAGGAGTTTGGGGAATATGATCAAGCAAAAGGCGACCAAGTTGCGGCGCATAT 660
QY 913 CATTAATCTGACAGCAGCAATACACAAACAGGCGGAGAGGCGGCTGCTCAACT 972
Db 661 CATTAATCTGACAGCAGCAATACACAAACAGGCGGAGAGTATGATATCTAT 720
QY 973 CTGAGCGGCAATGCGGCAACGAGGCGCAATACGAGGCGCATTCCTATTCAGGCTCAAGC 1032
Db 721 TTGGAGAGGCGATTTGCTGTAAGGAGGAGATATGCTCATTACGATATGCGGCTCAAG 780
QY 1033 GCGCAGAGGCGTTCGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1092
Db 781 GGGGACAGTGTCTTCCGATGTTTATTTATTTATTTATTTATTTATTTATTTATTTAT 840
QY 1093 GCGCTATTTAGAGACCGGCAACCTTGGGCGGAGAGATACATTCACATGCTATACGC 1152
Db 841 GGGATATTTAGAGAGCGCAACCTTGTGAAGGCAAGAAATGGGTTTCAATTTGCTGCGC 900
QY 1153 AAGCTTTTATTTATGAAATCTCTGAAAGATTTGCTATACGTTTATGAGCCATCG 1212
Db 901 AATCTTATTTTATGAAATTTTCGAAAGATTTTACATACATCTTACACCCGAGCT 960
QY 1213 GCGCATGCTCATACACATCATACAGACAAAGCGGAGCGGATGTCGAAACAAACA 1272
Db 961 GGTATGAGAGTGTACCATATTTAGTGAATGATATGTCAGGAGGCTATATACCTGAGAA 1020
QY 1273 ACAGAGACCATCTGAAGTCCGATCGGTTTAAAGACGACAAATTAACGCGGAAGT 1332
Db 1021 TCAGAGATACCATCAAAATTAATTAAGTACGTAAGCAATATAGTTTACCTTTAAAGAG 1080
QY 1333 AAAGACGATTTTACCA-----ATACCAAGGCGCAATATATTAACCTCCGCTTGAAT 1386
Db 1081 AAGATTAAGTTCTAATCTTACATATTTGACGACACTAATATTTATCTCCACGCTTAAAC 1140
QY 1387 AACGCTGAAACCTGTATTTTGGAGATCAAAAACGCGCATGTTACTTATCAACCAAC 1446
Db 1441 AATGAGAAACCTGTATTTTATGATGATCAAAAACAGAGATCTTATTTGATGATCGAC 1200
QY 1447 ATCAACCAAGGCTGGGCTGTTGATTTTGGAGGTAATCTTACGATCTTCAAGAAAT 1506
Db 1201 ATTAACCAAGGCTGGGCTGTTTATTTTGGAGGTAATTTTACGATCTTCAAGAAATCT 1260
QY 1507 AATGCACTTGGCAGAGTCTGAGAGTCAATGATGATGAGAGACAGTACTTCTGAGAA 1566
Db 1261 AACCAACTTGGCAGAGTCTGAGATCAATGATGATGAGAAATGACACCTGCTGAGAA 1320

| | | | |
|----|------|--|------|
| QY | 2829 | ACGTGGATTAATGCAAACTGGAACGGTATATGCAATTTAAGGAGTGGCTGCAATTTCT | 2868 |
| Db | 2401 | ACGGGTATTAATGCAAACTGGAACGGTATATGCAATTTAAGGATTTAGCTCAATTTTCT | 2460 |
| QY | 2889 | TTAAAAACAGCCATTTTTGGACCAAAATTGAGGGCGACAAGACACAAACGTAGCGTTG | 2748 |
| Db | 2461 | TTAAAAACAGCCATTTTTGGACCAAAATTGAGGGAGACAAGACACAAACGTAGCGTTG | 2520 |
| QY | 2749 | GAAATGCGAGCTGAGCAATGGCTAGCGATATGATATGAGAAATTTTAACGCTTAATAT | 2808 |
| Db | 2521 | GAAATGCGAGCTGAGCAATGGCTAGCGATATGATATGAGAAATTTTAACGCTTAATATAC | 2580 |
| QY | 2809 | AGTACTGTTACGTTAAATTCAGCTTATTCAGCTACTCAAAATATGCGCACGTACCGC | 2868 |
| Db | 2581 | AGTACTGTTACGTTAAATTCAGCTTATTCAGCTACTCAAAATATGCGCACGTACCGC | 2637 |
| QY | 2869 | CGTTCATTAAGACGGGAAACAAGCCCAATCGTGAAGAAACATCGTTTAAACAATTGACA | 2928 |
| Db | 2638 | CGTTCATTAAGACGGGAAACAAGCCCAATCGTGAAGAAACATCGTTTAAACAATTGACA | 2637 |
| QY | 2929 | GTAATGTTAAATTAGACGGGGCAGGACCATTCGAATTTACTTCACTTTATTTGGCTAT | 2988 |
| Db | 2698 | GTAATGTTAAATTAGATGGGGCAGGACCATTCGAATTTACTTCACTTTATTTGGCTAT | 2757 |
| QY | 2989 | AAAGCATTAATTAATATATCTATATGACGCTGAAGCGGATACATTAAGCTGTGGC | 3048 |
| Db | 2758 | AAAGCATTAATTAATATATCTATATGACGCTGAAGCGGATACATTAATATCTGTGGC | 2817 |
| QY | 3049 | GACACAGGCAAAAGACTGTGACCCCTTGAGCAATTTACTTTATTTGAAGCTTGATAT | 3108 |
| Db | 2818 | AACACAGGCAAAAGACCCCGAAACCTTGAGCAATTTACTTTGTTGAAGCAAAAGATAT | 2877 |
| QY | 3109 | CAACCCCTGGCAGTATAGCTTAAATTAATTTAATTTAAAAATTAACAGCTTGATGCGGTGCA | 3168 |
| Db | 2878 | CAACCGTTTACAGTATAGTCAATTTACTTTAGAAAATGACACAGCTTGATGCAAGTGA | 2937 |
| QY | 3169 | TGGCGTTATTAATTAGAGAAAGAAAAGCGGAAATTCGGCTTCATATTCCAATTAAG | 3228 |
| Db | 2938 | TTAGCTTATTAATTAGAGAAATGATGGGAAATTCGGCTTCATATTCCAATTAAG | 2997 |
| QY | 3229 | CAGGAATTTGGCAATGATTTAGTAAAGACAGACAGTAAAGACGATTTAGAAAGCAAA | 3288 |
| Db | 2998 | CAGGAATTTGCACATGATTTAGTAAAGACAGACAGACGAAACGAACTTTAGAACCAA | 3057 |
| QY | 3289 | CAACCTGAACGACTACTATAAAACAAAACCTGAGAGCTAAAGTCGGTCAAAA----- | 3342 |
| Db | 3058 | CAAGTGAACGACTGCTATAAACCAACACAGGAGACCCAAAGTGGGTCAAGAAAGACA | 3117 |
| QY | 3343 | --AGAGGGCGGTTTCTGATACCCCGCTGATCAAAAGCCAGTTAAACGATTTACAAACC | 3399 |
| Db | 3118 | GCGAGAGAGCGGTTTCTGATACCCCGCTGATCAAAAGCCGTTAAACGATTTAAAGCC | 3177 |
| QY | 3400 | GAACTCGAGAGATTAATGCCCACAAAGCAAGTGGCAAGCGGTGCAAAATACAGAAATGA | 3459 |
| Db | 3178 | AAACAGCTGAACGACTGCTGTAACACAA-----AAAAGTAAAGCAAAAACAAAAAAG | 3222 |
| QY | 3460 | ACTGCATTTAACCAAAAGAACGACAAAGTTAAACCACTCAAGATTAACCAAAATTTAGTC | 3519 |
| Db | 3233 | TGGCGTCAAAAAGACAGCTGTTTTC-----TGATCCCTCTTGATCAAAAGCCGTGGTCCA | 3288 |
| QY | 3520 | TTGGCACTGCATTTGGAGGAAAAAGAACCGCTCAGATTTGATTTGCTAATCAAAATTA | 3579 |
| Db | 3289 | TTAGAGACCGGCACTTGAGGTTATTTGATGCCCCAC----- | 3322 |
| QY | 3580 | GCTCAGTTGAATTTAACACACAACTTAGAAAAAGCCTTAGCAGTGGCTGAGCAAGACAGA | 3639 |
| Db | 3323 | -----AGCAATCGGAA | 3333 |
| QY | 3640 | AAAGAGCTTAAGCTCAAGAGACAGCAAAAGACAAAGCAAAACAAACACTTGAATGACG | 3699 |
| Db | 3334 | AAAGATTCCTCTAGCTCAAGAAAGAGCGGAAAAACAAACGCAACAAAAAATCTTGATGACG | 3393 |
| QY | 3700 | CGTTATTCAAATAGTCGTTATTCAGAAATTAATTCGAACAAGTAATAGTATGCTTCCGTT | 3759 |

Wed May 7 15:00:47 2003

pct-us03-05226-16.rge

Page 4

pct-us03-00

240
ה

14. *Journal of the American Medical Association*, 1997; 278: 1039-1044.

| | | | |
|----|------|---|------|
| Db | 3394 | CGTATTCAAATAGTCCGTTATCAGAAATATCTGCACACGATAAATAGATGCTTCTGTT | 3453 |
| Oy | 3760 | CAGATGAATATATATGCTTTTGTATATCAACTCAATCTCGGAGGGAACAATATC | 3819 |
| Db | 3454 | CAGATCAATATGATGCTTTTGTATATCAAGCAAAATCTGCTGTGGACAAATATC | 3513 |
| Oy | 3820 | TCACAGATTAAGACGTTATGATTCTGATGCGTTCGTCCTATACGCAAGAAAGAAC | 3879 |
| Db | 3514 | GCACAGGATTAAGACGCTATGATCTGATCGTTCCGTCCTATACGCAAGAAAGAAC | 3573 |
| Oy | 3880 | TTGGCTCAATTTGGGTTGGCAAAAACCTTAGCTATACGAGAGAAATTTGGAGTTTCTG | 3939 |
| Db | 3574 | TTAGCTCAATTTGGGTTGGCAAAAACCTTAGCTATATGACGAATTTGGGCAATTTTCTG | 3633 |
| Oy | 3940 | CATAGCCGTTACAGATATATCTTTGATATACAGGTTAAATATCAGCAACATTAAAGATG | 3999 |
| Db | 3634 | CATAGCCGTTACAGATATATCTTTGATATACAGGTTAAATATCAGCGACATTAAGATG | 3693 |
| Oy | 4000 | ATGCGGGTTTCCCAATATCAATGAGGCGTATTTACAAATTTGGTGTAAACGCGGAAG | 4059 |
| Db | 3694 | ATGCGGGTTTCCCAATATCAATGAGGCGATTTACAAATTTGGTGTAAACGCGGAAG | 3753 |
| Oy | 4060 | GGAATTTGTCCGAGTAAATGGCTGAAGACAAAGCCGAAAATTCATCGAAAAGGATAT | 4119 |
| Db | 3754 | GGAATTCAGTCCGATTAATGGCTGAAGACAAAGCCGAAAATTCATCGAAAAGGATAT | 3813 |
| Oy | 4120 | AATATGCGCTGATCCAAAGTATTCGTTCCATTTAGGCGCAATGCGTATTCAGCCTAT | 4179 |
| Db | 3814 | AATATGCGCTGATCCAAAGTATTCAGTTCCGTTTAGGCGCAATGCGCATTCAGCCTAT | 3873 |
| Oy | 4180 | TTTGAGATTAATGCTATTTTATGAAGTAAATTTCAATCTGAGGAAGTAAAGT | 4239 |
| Db | 3874 | TTTGAGATTAATGCTATTTTATGAAGTAAATTTCAATCTGAGGAAGTAAAGT | 3933 |
| Oy | 4240 | CAAAACCGAGCCTGCATTTAAATGCTTAATGCTGAGTAGGGTGCATTAACGTT | 4299 |
| Db | 3934 | AAAACCGCTACGCTGATTTAATGCTTAATGCTGACATTCAGATTTGATTAATTT | 3993 |
| Oy | 4300 | ACCCGACAGAGATATACGCGTTAAGCCTTTATTTCTTCGTCATATATGTTGATGTTCA | 4359 |
| Db | 3994 | ACTCGACAGATATATACGCGTTAAGCCTTTATTTCTTCGTCATATATGTTGATGTTCA | 4053 |
| Oy | 4360 | AACGCTACAGTCAAAACCCCTGTAATCCGCGGCTGTCGAACAACATTTGACGTTAT | 4419 |
| Db | 4054 | AACGCTACAGTCAAAACCCGCGTAATCTCAGGCTGTGCAACAACATTTGACGTTAT | 4113 |
| Oy | 4420 | TGGCAAAAAGAGTGGATTAAAGCGGAATTTTCAATTTCAACTTTTGCCTTTAT | 4479 |
| Db | 4114 | TGGCAAAAAGAGTGGATTAAAGCGGAATTTTCAATTTCAACTTTTGCCTTTAT | 4173 |
| Oy | 4480 | TCTTAATCTCAAGTTTGGCAATCGGTTAAACGCAATATGCGCGTAATTAGATAT | 4539 |
| Db | 4174 | TCAAAATCTCAAGTTTACAACTCGGCAAAACGCAAAATTTGGGCGTGAATTTGGCTAT | 4233 |
| Oy | 4540 | CGTTGTGTAATAATCAATATTTTAT | 4566 |
| Db | 4234 | CGTTGTGTAATAATCAATATTTTAT | 4260 |

| | | | | | |
|------------|------------------------------------|---------|-----|--------|-----------------|
| RESULT 2 | | | | | |
| AR157320 | | | | | |
| LOCUS | AR157320 | 4319 bp | DNA | linear | PAT 17-OCT-2001 |
| DEFINITION | Sequence 1 from patent US 6245337. | | | | |
| ACCESSION | AR157320 | | | | |
| VERSION | AR157320.1 | | | | |
| KEYWORDS | GI:16218252 | | | | |
| SOURCE | Unknown. | | | | |
| ORGANISM | Unknown. | | | | |
| REMARKS | Unclassified. | | | | |
| | Accession bases 1 to 4319) | | | | |
| | W. W. III and Falkow, S. | | | | |
| | adherence and penetration proteins | | | | |

| | | |
|----------------------------|---------------------------|--------------------------------|
| JOURNAL | Patent: | US 6245337-A 1 12-JUN-2001; |
| FEATURES | Location/Qualifiers | 1..4319 |
| SOURCE | /organism="unknown" | |
| BASE COUNT | 1497 a 776 c 891 g 1155 t | |
| ORIGIN | | |
| Query Match | 61.5%; | Score 2970; DB 6; Length 4319; |
| Best Local Similarity | 82.3%; | Ped. No. 0; |
| Matches 3581; Conservative | 0; Mismatches 645; | Indels 124; Gaps 10 |

QY 253 TCAATAGTCGTTAACCAGCAGTATTTTAAATACGAAAAATTAAGTAAATTAATTAACATT 312
 Db 1 TCAATAGTCGTTAAGTA -GTATTTTAAATACGAAAAATTAAGTAAATTAATTAACATT 359
 QY 313 ATGAAAAAACTGATTTGCTGCAATTTTAAACCGGTGCATTCAATGAGGATGTA 372
 Db 60 ATGAAAAAACTGATTTGCTGCTAATTTTAAACCGGTGCATTTCATTGGGATAGTA 119
 QY 373 TCGCAGCGTGGCGAGCTCATCTATTTTGGGATTAACACCAATATTAATCGTATTT 432
 Db 120 TCGCAGCGTGGCGGTGCTGCACACTTTTGGGATTAATACCAATATTAACGATTT 179
 QY 433 GCCGAGATTAAGGAGATGCACAGTTGGGGCTAAATAATATGAGGTTACATTAATAAT 492
 Db 180 GCCGAGATTAAGGAGAGTTCACAGTTGGGGCTCAAAATATTAAGGTTATTAACAAACA 239
 QY 493 GGAATTTAGTGGCACAATATGACAAAGGCCCAATGATGATTTTCCGTGCTGTCG 552
 Db 240 GGGCATTAGTTGGCACAATATGACAAAGGCCCGCATGATTTTTCGTAGTGCA 299
 QY 553 CGAATGGGAGTGGCGCATGCTGGCGCATACGATATTTGTAGTGGCACAATATGTA 612
 Db 300 CGTAACGGCGGCGCAGCCTTGTGTGAATAATATATGTAGCGGTGGCACAATAGCTA 359
 QY 613 GGCATACCAATGCGAATTTGGTGGTGAAGGACAAATCCGTGATCAGATCGTTTACT 672
 Db 360 GGAATACAGATGTGATTTTGGTGGAGAGGAAACAAACCCGATCAACATCGTTTACT 419
 QY 673 TATTAATTTGGAACGCAATATTTATTAACGATCAACACCATCTATGAGAAAGAC 732
 Db 420 TATTAAGATTGTAACGCAATTAACCTCAAAAAGATTAATTACATCTTATAGAGCAGAT 479
 QY 733 TACCAACAACCCAGCTTACATTAATTTGTTACGGAAAGCCACCCCAATGATGACTTCT 792
 Db 480 TACCATTAATCCACGATTAACATTAATTCGTTACGAAAGCGGCTCCCAATGATGACTTCG 539
 QY 793 GATTAAGAACGGCACAATATACAGATAGAGCAAAATCCCGAAGCGGTGATCGGC 852
 Db 540 AATATGAAATGCGCACTACTTATTCAGATAGAACAAATATCCGAACGTTGCTATCGGC 599
 QY 853 TCCGGGTGCGATTTTGGCGAAACGATCAAAACACGCGACCAAGTTGCGCGCGCATAT 912
 Db 600 TCTGGACGCGAGTTTGGCGCAAAATGATCAAGAACGAGGACCAAGTTTCCGGTGCAGAT 659
 QY 913 CATTAACCTGACAGCAGCATATACACAAACCAAGCGGAGCAGCGAGGGGGCGGTGCTCAGT 972
 Db 660 CATTAATCTGACAGCTGGCGAATACACAAATCAGCGGAGCAGGTAAATGATATTCGAT 719
 QY 973 CTGACAGCGCATGTCGCGCAAGGGGCAATTAAGCGCCCATCTCATGACAGGCTCAAGC 1032
 Db 720 TTGGAGCGCGATGTTCTGTAAGGCGGGAGATATGCTCATTCACCATTTGCGAGCTCAAG 779
 QY 1033 GCGCAGCAGCGGTTGCGCTATGTTATTTATATGATGCGGAAAAACAATAATGTTGTTAAAC 1092
 Db 780 GGGGACAGCTGTTCTCCGATGTTATTTATATGATGCGGAAAAACAATAATGTTTAAAT 839
 QY 1093 GGCCTATTGAGAGACCGGCAACCCCTGGGCGGGGACAGATAACATTCCACTGGTATAC 1152
 Db 840 GGGATATTACGGGAAGCCACACCCTTTGAAGGCAAGAAATAGGTTTCATATGTTTCCG 899
 QY 1153 AAGCTTTTTTGATGAATTCCTGAAAAGATTTGCGCAATTCGCGATTTGCGCGATTT

```

OY 3343 ---AGAGGGGGTTTTCTTGATACCCCGCGCTGATCAAAAGCCAGTTAAAGCATTTACAGCC 3399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3117 GCGAGAGCAGCGTTTCCGATACCTTGCCTCATCAAAAGCCGTTAAAGCATTTGAAAGCC 3176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3400 GAAGTCGAGAGCAATTAATGCGCAACAGCAAGTGGCACAAGCGGTGCAAAATCAGAAAGTA 3459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3177 AAACAGAGCTGAAGTACGCTGTAACACAA-----AAAGTAAGCAAAACAAAAAG 3231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3460 AOTGCACTTAACCAAAAGAGCAAGCAAGTAAACCACTCAAGTAAAGCAAAATTTGATC 3519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3232 TGGCGTCAAAAGAGCAAGTCTTTTC-----TGATCCCGCTGATCAAAAGCGCTTGGCA 3287
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3520 TTGGCACTGATGTTGTAAGAAAGAAACCGCTCAGATTGATTTTGCTAATGCAAAATTA 3579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3288 TTAGAAGCCGCACTTGAGATTGATGATGCCCCAC----- 3321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3580 GCTCAGTTGAATTTACACAAACACTAGAAAAAGCCTTAGCAGTGGCTGAGCAAGCAGAA 3639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3322 -----ACCAATCGGAA 3332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3640 AAAGAGCGTAAAGCTCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3333 AAAGATCGTCAAGCTCAAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3332
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3700 CGTATTCAAATAGTGGCTATACAGAAATTAATCTGCACAGTAAATAGTATGCTTTCGCTT 3759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3393 CGTATTCAAATAGTGGCTATACAGAAATTAATCTGCACAGTAAATAGTATGCTTTCGCTT 3452
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3760 CAAGATGAATTAATGATCGCTTTTGTAGTCAAGTCAATCGCGTGGCAAAATATC 3819
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3453 CAAGATGAATTAATGATCGCTTTTGTAGTCAAGTCAATCGCGTGGCAAAATATC 3512
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3820 TCACAGATAAAGACGTTTATGATTCGATGCGTTCCTGCTAT--CAGAGAAAAG 3876
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3513 GCACAGGATAAAGACGTTTATGATTCGATGCGTTCCTGCTAT--CAGAGAAAAG 3572
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3877 AACTGCGTCAATGGGGTGGCAAAAAGCCTTAGTCAAGCAAGCAAGTGGGCAATTTTC 3936
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3573 AACTGCGTCAATGGGGTGGCAAAAAGCCTTAGTCAAGCAAGCAAGTGGGCAATTTTC 3632
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3937 TGGCATAGCCGTTCAAGTATATCTTTTGTATGACAGGTTAAATCAGCAATTAACG 3996
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3633 TGGCATAGCCGTTCAAGTATATCTTTTGTATGACAGGTTAAATCAGCAATTAACG 3692
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 3997 ATGAGTGGGTTTGGCCCAATATGATGAGTGGGTTTACATTTGGTGAATGAAGTGGGA 4056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3693 ATGAGTGGGTTTGGCCCAATATGATGAGTGGGTTTACATTTGGTGAATGAAGTGGGA 3752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4057 ACGGGAATTAATGCGAGTAAATGGCTGAAGAAAGCAAGGCGAAATTCATCGAAAAGCG 4116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3753 ACGGGAATTAATGCGAGTAAATGGCTGAAGAAAGCAAGGCGAAATTCATCGAAAAGCG 3812
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4117 AATAATTAATGCGTGAATCAATTAATGCTTTCATTTAGGCAATTTGGGCTTTAGCGCT 4176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3813 AATAATTAATGCGTGAATCAATTAATGCTTTCATTTAGGCAATTTGGGCTTTAGCGCT 3872
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4177 TATTTGAGTAAATGCTATTTTATTAAGCTAAATTAATCAATCTGAGCAAGTGAAG 4236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3873 TATTTGAGTAAATGCTATTTTATTAAGCTAAATTAATCAATCTGAGCAAGTGAAG 3932
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4237 GTGCAACACCGAGCCTTGCAATTAATGCTGATTAATGCTGAGTACGATTAATGAG 4296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3933 GTGCAACACCGAGCCTTGCAATTAATGCTGATTAATGCTGAGTACGATTAATGAG 3992
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4297 TTACCCGAGAGAGATTAATGAGCGTTAGCCCTTATTTCTGTCATTAATTTGATGTT 4356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3993 TTACCCGAGAGAGATTAATGAGCGTTAGCCCTTATTTCTGTCATTAATTTGATGTT 4052
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4357 TCAACGCTAGCTACAAACCACTGTAATTCGCGCGGTGTGCAACACCAATTTGAGCT 4416
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4357 TCAACGCTAGCTACAAACCACTGTAATTCGCGCGGTGTGCAACACCAATTTGAGCT 4112
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 4113 TATTGGCAAAAGAGTGGGATTTAAAGCAGAAATTTTCAATTTCCAAATTTCCGCTTT 4172
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4477 ATTTCTAAATCTCAGGTTCCGACCTCGTTAAACAGCCGAAATTTGGCGCTGAATTAAGA 4536
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4173 ATCTCAAAATCTCCAGGTTCCACAACTCGGCAACCAAAATGTTGGCGTGAATTTGGCG 4232
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 4537 TATCGTTGGTAAATCAATCAATTAATTTAT 4566
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4233 TATCGTTGGTAAATCAATCAATTAATTTAT 4262
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
U32710/C
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
1 (bases 1 to 13818)
Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
Merriam,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
Gocayne,J.D., Scott,J.D., Shiley,R., Liu,L.I., Glodek,A.,
Kellay,J.M., Weidman,J.F., Phillips,C.A., Spitzgs,T., Hedblom,E.,
Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhmann,J.L.,
Georgagen,R.S., Gnehm,C.L., McDonald,D.A., Small,K.V., Fraser,C.M.,
Smith,H.O. and Venter,J.C.
Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd
Science 269 (5223), 496-512 (1995)

JOURNAL
MEDLINE
PUBMED
7542800
2 (bases 1 to 13818)
Tatusov,R.L., Mshagian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
Borodovsky,M., Ruid,K.E. and Koonin,E.V.
Metabolism and evolution of Haemophilus influenzae deduced from a
whole-genome comparison with Escherichia coli
Curr. Biol. 6 (3), 279-291 (1996)
96398784
MEDLINE
PUBMED
8805245
3 (bases 1 to 13818)
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
Direct Submission
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
4 (bases 1 to 13818)
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
Direct Submission
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
The H. influenzae sequence has been updated by R. Fleischmann. New
database matches have been assigned, product names have been
improved, and a number of frame shifts have been corrected. We
gratefully acknowledge the work of Tatusov et. al. We have
incorporated their annotation into the /notes fields of the
corresponding H. influenzae genes
5 (bases 1 to 13818)
White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
Direct Submission
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
The whole genome was shifted by 588 nucleotides for a new start
on Sep 30, 1996 this sequence version replaced gi:1220911.
Location/Qualifiers
1. 13818
/organism="Haemophilus influenzae Rd"

```

*TGGGATTAAGGGAATTTTCAATTTCCAACTTTCTGCTTT 4476

| | | | |
|----|-------|---|-------|
| QY | 125 | AACTATTAAAGCAAAAACCCAGAAATTTGGCTTAATACATATAGTTTACATCATTTT | 184 |
| Db | 13818 | AAAGTATTAAAGCCAAAACCTAGAAATTTGGCTTAATACATATAGTTTACGCTTTAT | 13759 |
| QY | 185 | TTTCTTTGTGGCTTTTACTTCGTTTTTTTAACTGGAATCCCTTAGAAATACGCGACT | 244 |
| Db | 13758 | TTTTCTTTGTGGCTTTTAACTTCG - TTTTAACTGGAATCCCTTAGAAATACGCGACT | 13700 |
| QY | 245 | TTTATTGTTCAAATAGTCGTTTAAACGCTATTTTAAATCGAAAAATTAATCTAATTA | 304 |
| Db | 13659 | TTTATTGTTCAAATAGTCGTTTAAACGCTATTTTAAATCGAAAAATTAATCTAATTA | 13640 |
| QY | 305 | TAAACATTTATGAAAAAAACTGTATTTGCTCGAATTTTTTAAACGCTTCGATTCATAG | 364 |
| Db | 13639 | TAAACATTTATGAAAAAAACTGTATTTGCTCGAATTTTTTAAACGCTTCGATTCATAG | 13580 |
| QY | 365 | GGATTAATATCGAAGCGGTGGGAGCTCAATCTATTTTGGATTTGACTACCAATATTAC | 424 |
| Db | 13579 | GGATTAATATCGAAGCGGTGGGAGCTCAATCTATTTTGGATTTGACTACCAATATTAC | 13520 |
| QY | 425 | GTTATTTTCCCGAGAAATAAGGAAAGTTCCACAGTTGGGCTAAAAATATTGAAGTTTAC | 484 |
| Db | 13519 | GTTATTTTCCCGAGAAATAAGGAAAGTTCCACAGTTGGGCTAAAAATATTGAAGTTTAT | 13460 |
| QY | 485 | ATTAATAATGGAATTTAGTTGGACATCAATAGCAAAAAGCCCAATGATGATTTTCCG | 544 |
| Db | 13459 | ATTAATAATGGAATTTAGTTGGACATCAATAGCAAAAAGCCCAATGATGATTTTCCG | 13400 |
| QY | 545 | TGGTGTCCGGAATTTGGGGTGGCGGCTTTGGTGGCGATCAGTATATTGTAGTGTGGCAC | 604 |
| Db | 13399 | TGGTGTCCGGAATTTGGGGTGGCGGCTTTGGTGGCGATCAGTATATTGTAGTGTGGCAC | 13340 |
| QY | 605 | ATAATGTAGGCTTTACCAATGTGATTTTGGTGTGGAAGGACAAAATCCGATCAACATC | 664 |
| Db | 13339 | ATAATGTAGGCTTTACCAATGTGATTTTGGTGTGGAAGGACAAAATCCGATCAACATC | 13280 |
| QY | 665 | GTTTACTATTAATAATTGTGAAAGCGAATATTATAA---AACGATCAAGCATCTTT | 721 |
| Db | 13279 | GTTTACTATTAATAATTGTGAAAGCGAATATTATAA---AACGATCAAGCATCTTT | 13220 |
| QY | 722 | ATGGAAGAAGCTACCAACACCGCTTACATAATTTGTATAGGAAGCCACCCCAATG | 781 |
| Db | 13219 | ATGATGGAATTTACATATGCTCGTTTACATTAATTTGTATACGAAAGCTGAAACCTGTG | 13160 |
| QY | 782 | ATATGACTTGTGATATGAAGCGCAACAAATATACATATGAGACGAATATCCGAACG | 841 |
| Db | 13159 | GTAATGACAAATAATGATGGAAGATATATGCTGATAGAGAAATATCTGGAACGTG | 13100 |
| QY | 842 | TGCGTATCGGCTCCGGGTGGAGTGTGGCGGAAGCATCAAAACGAG----- | 889 |
| Db | 13099 | TACGTATAGGCTGAGACGTCACTATTGGCTACAGATTAAGATGAAGAAACGATGTAC | 13040 |
| QY | 890 | -----GCGACCAAGTTGCCGGCGCATATCATTAAGCTGACAGCGACGCAATACACACAC | 943 |
| Db | 13039 | ATAATTCATATTATGCTCAGGTGATATCGTATCTTACTGCGAGGAATATCCCATATC | 12980 |
| QY | 944 | AAGCGGAGCAGGAGGCGGCTGGTCAATCTGAGCGGGGAGTGTGGCGCAAGCGCGCAAT | 1003 |
| Db | 12979 | AGAGTGAATATGTAATGTAAGTCATATCTTAATGGTAATGTATGTTAGCCCTAATCAT | 12920 |
| QY | 1004 | ACGGCCCATTTCTATTCAGGCTCAACGCGGACAGCGGTTGCGCTATGTTATTATG | 1063 |
| Db | 12919 | ATGTCCTATATCCACAGCGGTGTTCTAAAGGCGATAGCGGTTGCCCATGTTATTATG | 12860 |
| QY | 1064 | ATGCGGAAAAACAAATGTTGATTAAGCGGATATGAGAACCGGCAACCTTTGGCGG | 1123 |
| Db | 12859 | ATGCGGAAAAACAAATGCTTATTAATGCTGATTTCAAACTGGGCACTCTTTTTCG | 12800 |
| QY | 1124 | GGACAGAAATACATTCACACTGTAGC---CAAGCTTTTATTGAGAAATCTTGAA | 1180 |
| Db | 12744 | GGACAGAAATACATTCACACTGTAGC---CAAGCTTTTATTGAGAAATCTTGAA | 12740 |

| | | | |
|---|-------|--|-------|
| D | 12739 | TTGATACCCCTACGTGTTTTTCACCGCTAATATTCGCCCAATAAAGSACATATATTCCTTGG | 12688 |
| Q | 1235 | CAGACAAAGCGCGAGCGGCGGCGATGTCA-----AACACAA | 1273 |
| D | 12679 | TATCAATATATGATGCTACAGGTAAATTACCTTAACTAGACCTTAGTAAGATGGCTCTA | 12622 |
| Q | 1274 | CAGSAGACCATCTGAAAGTCCGATCGTTTAAAGACGACAAATTACCTCCGAAAGTA | 1333 |
| D | 12619 | AAGCAAAATCAGAGTAGGAACGTGTAAGTTATTTAAATCCATCGTTAAATCAAAACCTTA | 12566 |
| Q | 1334 | AAGCGATGTTTACCAATACAAAGGCCAAATATATACCTCGCTTCGTTGAATTAAGCTG | 1393 |
| D | 12559 | AAGAACATGTTAAAGCAGACGAGCGGTATATATTTACCAACCAAGATGGAATATGAA | 12500 |
| Q | 1394 | GAACCTGTATTTCCGAGATTCACAAAAAGCGACCTGTACCTTATCAACCAATCAACC | 1453 |
| D | 12459 | AAATATTTTACCTTGGCGACCAAGGAAAGAACTTTAACTATCGAAATATATTAATC | 12444 |
| Q | 1454 | AAGGTGGGGCGGTTGTATTTTGGAGGTAACCTTAACGTT---ATCTTCAAAAATATAG | 1510 |
| D | 12439 | AAGGTGCTGTGATATATCTTGAAGGTAATTTTGGTTAAAGGCAAGCAAAATATATA | 12388 |
| Q | 1511 | CAACTTGGCAGAGTGCCTGAGTGCATAGTAGTGAAGACAGTACTGTACTTGGAAAGTAA | 1570 |
| D | 12379 | TAACTTGGCAGAGTGCAGGCGGTATCTATTTGACAAAGTACCACTGTTGAATGAAAGTTC | 12322 |
| Q | 1571 | ATGGTGTGAAATGATAGCGCTTCTTAAATGGCAAGCGCAATGACGTTAAACCA | 1630 |
| D | 12319 | ACAAATCTGAAATGATCGTTTATCTTAAATATGTAAGTACCACTTATTAATCATGTGA | 12266 |
| Q | 1631 | AAGGGGAAATTAAGGTTTCATCGATCGCTAGCGGATGTAAGTCAATTTTGGAGCAGAGG | 1690 |
| D | 12259 | AGGAAAGAAATTTAGGAAGTTTAAGTGGCGGTAACGGCAAGTCACTTACATCAACAG | 1220 |
| Q | 1691 | CAGACGATCAGGCACAAACAAGCCTTTAGTGAATAGGCTGGTTAGTGGCAGAGTA | 1750 |
| D | 12199 | CAGATGAAGCGGTCACAAAACAAGCTTTAAAGAAAGTGGCATGTGAAGCGGTGCACAA | 12144 |
| Q | 1751 | CGGTTCAGTTAAACGATGACAAACAATTAATCTAGTAATTTATTTCCGCTCCGTG | 1810 |
| D | 12139 | CAGTTCAATTAATAGTACGATCAAGTATGATCTTAACAATATCTATTTGGATTTGCTG | 1208 |
| Q | 1811 | GTGTCCTTATGATCTTAATGGGCAATTCATTAACCTTTAAACGTAATCCAAATACGATG | 1870 |
| D | 12079 | GTGTCCTTATGATCTTAAAGGGCAATTCATTAACCTTTAAAGGTAATCCAAATACGAGC | 1202 |
| Q | 1871 | AGGAGACAGATGTTTAATCACAATGACCAACAGAAATCTACAGTACGATTAACCTGCA | 1930 |
| D | 12019 | AGGCGCGATGATGTGAAACATTAATACATCACTCAAGTGGCTAATTTACTTCTGGGA | 1196 |
| Q | 1931 | GGATACCATTAATG-----ACACACTGGCGATTAAACAATAACGTG | 1975 |
| D | 11959 | ACGAAATATTACTGCTCCATCTAATAAAATAATTAATTAACCTGATTAACGACAAAG | 1190 |
| Q | 1976 | ATATGCTTTAATGATGGTGTGATGTAAGATGATTAATTAATTAATTAATTAATTAATTAAT | 2035 |
| D | 11899 | AAATGCTTACAAAGCTGTGTTGGCGAAACGATTAATTAATTAATTAATTAATTAATTAAT | 11844 |
| Q | 2036 | ATGTTACTTAACAATCGCTTAACAAGATTAATCACTTCTCTTATCAAGTGAACAATT | 2095 |
| D | 11839 | ACCTTATTTTAAACCAACACAGAGATCGTACTTGTACTTCAAGTGGCAAACT | 11788 |
| Q | 2096 | TAAAGGCAATTAATCAAGAGGCGCATTTAGTGTGTTAGTGGTGGCCACACAC | 2155 |
| D | 11779 | TAAAGGCGATTAATCAACAAAGGTAACCTTTTTCAGCGGTAGCCGACACCC | 11722 |
| Q | 2156 | ACGATACATCAATTTAA-----TCGCTTAAAGAGATGTTGGCGACCTTAAGGCGAGAG | 2209 |
| D | 11719 | ACGCTCATCATTTTAGCAAAACGTTGCTGAAATAGGAAGTATCCACAAAGCGCAA | 1166 |
| Q | 2210 | TGCTTATTAATGATGAGATGGAATCAACCGTATTTAAAGCTGAAAACTTCCAAATTAAG | 2269 |

SANCTI

2210 TGGTATTGATGACGATTGGATCAACCGTACATTAAAGCTGAACAATTAAAG 2269

Med May 7 15:00:47 2003

pct-us03-05226-16.rge

Page 9

Db 11659 TTGTGTGGGATTTAGCATTCACCGTACATTTAAAGCTGAAGAACTTCCAAATTTAAG 11600
QY 2270 GCGGAGTACGGTGGTTCTCGCANGTTTCTTCAATTGAAGAAATGACATACGA 2329
Db 11599 GCGGAGTACGGTGGTTCTCGCANGTTTCTTCAATTGAAGAAATGACATACGA 11540
QY 2230 ATACGCGCAAGCGACATTTGGTGTGGCAATCAAGAAATACCATTTGACGGGT 2389
Db 11539 ATATGCGAAATGCCACATTTGGTGTGGCAATTTCAAAATACCATTTGACGGGT 11480
QY 2230 CAGATTTGAGAGGATTTAGCACTTGTAAACAGTTAAATTTAACCGATTAATG 2449
Db 11479 CAGATTTGAGAGGATTTAGCACTTGTAAACAGTTAAATTTAACCGATTAATG 11420
QY 2450 ATTCCATACGACACACAAATTAATGGCTCTATTTAATTTAATTAATGACAACTGA 2509
Db 11419 ATTCCATACGACACACAAATTAATGGCTCTATTTAATTTAATTAATGACAACTGA 11360
QY 2510 ATATTCATGCTTTAGCAAACTAATGATGTACCTTTAATTAATTAATGACAACTGA 2569
Db 11359 ATATTCATGCTTTAGCAAACTAATGATGTACCTTTAATTAATTAATGACAACTGA 11300
QY 2570 CATTGAGCACATGCGACCCAAACAGGCAATTCACATTTCAATTCAGCAATGCAA 2629
Db 11299 CATTGAGCACATGCGACCCAAACAGGCAATTCACATTTCAATTCAGCAATGCAA 11240
QY 2630 CCGTGGATTAATGCAAACTTGAAGGTTAATGCTAATTTAAGGATTTGCTCAATTTCT 2689
Db 11239 CCGTGGATTAATGCAAACTTGAAGGTTAATGCTAATTTAAGGATTTGCTCAATTTCT 11180
QY 2680 TAAAAACAGCCATTTTGGGACCCAAATTCAGGGGCAACAAACACAACTGACGTGG 11120
Db 11179 TAAAAACAGCCATTTTGGGACCCAAATTCAGGGGCAACAAACACAACTGACGTGG 11120
QY 2750 AAAAGGCACTTGGACATGCTAGCTACGATACATTCGCAATTTAAGCTTAATTAATA 2809
Db 11119 AAAAGGCACTTGGACATGCTAGCTACGATACATTCGCAATTTAAGCTTAATTAATA 11060
QY 2810 GTACTGTACGTTAATTAATGAGTTATTCAGTACGTTCAATTAATGCGCACGCGCC 2869
Db 11059 GTACTGTACGTTAATTAATGAGTTATTCAGTACGTTCAATTAATGCGCACGCGCC 11000
QY 2870 GTTCTTTGAGAGCGGAAACAAAGCCACATGCGAAGAAACATGCTTCAACATGACAG 2929
Db 10999 GTTCTTTGAGAGCGGAAACAAAGCCACATGCGAAGAAACATGCTTCAACATGACAG 10940
QY 2930 TAAATGTAAATTTGAGCGGCGAAGGACATTCGCAATTTACTTCATCTTTATTTGGCTATA 2989
Db 10939 TAAATGTAAATTTGAGCGGCGAAGGACATTCGCAATTTACTTCATCTTTATTTGGCTATA 10880
QY 2990 AAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3049
Db 10879 AAAGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10820
QY 3050 ACACAGGCAAAAGAACTGTGACCTTGACCAATTTACTTTAATTTAAGGCTTGATATC 3109
Db 10819 ACACAGGCAAAAGAACTGTGACCTTGACCAATTTACTTTAATTTAAGGCTTGATATC 10760
QY 3110 AACCTTGCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3169
Db 10759 AACCTTGCACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10700
QY 3170 GCGGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3229
Db 10699 TACGTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10640
QY 3230 AGGATTTGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3289
Db 10639 AGGATTTGCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10580
QY 3290 AAAGTAACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3349
Db 10579 AAAGTAACTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10520

QY 3350 CGTTTTCATATACCCCGCTGTATCAAGGCACTTAAGCCATTACAGGCCAATCGAGA 3409
Db 10519 TGTGTCTATATACCCCGCTGTATCAAGGCACTTAAGCCATTACAGGCCAATCGAGA 10465
QY 3410 CGATTATGCCCAACAGCAAGTGGGACAAAGCCGTGCAAAATCGAAGTAACTGCACTTA 3469
Db 10464 AGTTGACAGACTACTGGAACACAAAGCAAGTAAAGCAAAAGAAAGGCGGTCAAA 10405
QY 3470 ACCAAAAGACAGCAAGTAAACCACTCAAGATTAAGCAAAATTTAGCTTGGCACTG 3529
Db 10404 AAGAGCATTTGAGTGA----- 10389
QY 3530 CATTGTGGAAGAAACCGCTCAGATTTGATTTGCTAATGCAAAATTAAGCTCAGTTGA 3589
Db 10388 -----GCGTTTTCATATACCCCGCTGTATCAAGGCC 10358
QY 3590 ATTTAACACACACTAGAAAAAGCCTTAGCAGTGGCTGACGACGACAAAAAGACGTA 3649
Db 10357 AGTTAAAGGATTCGAAAGTCAAACTTGAGGTTATTAATCCCAACGCAAGTAAAAAG 10298
QY 3650 AAGCTAAGGCAAGGAAAGAAAGCAACGCAAAACAAAGACTGATCAGCCGTTATTCAA 3709
Db 10297 AAGCTAAGGCAAGGAAAGAAAGCAACGCAAAACAAAGACTGATCAGCCGTTATTCAA 10241
QY 3710 ATAGTGGTTATCAGAAATTAATTCGCAACAGTAAATAGTATGCTTCCGTTCAAGATGAAT 3769
Db 10240 ATAGTGGTTATCAGAAATTAATTCGCAACAGTAAATAGTATGCTTCCGTTCAAGATGAAT 10181
QY 3770 TAAATGCTCTTTTGTAGATCAGTCAATTCGCGGTGGGACAAATTAATTCAGAGATA 3829
Db 10180 TAAATGCTCTTTTGTAGATCAGTCAATTCGCGGTGGGACAAATTAATTCAGAGATA 10121
QY 3830 AAAGGCTTAATGATTCGATGCGTTCGCTTATCAGAGAAAGCAACTTGCCTCAA 3889
Db 10120 AAAGGCTTAATGATTCGATGCGTTCGCTTATCAGAGAAAGCAACTTGCCTCAA 10061
QY 3890 TTGGGCTGCAAAAGCCCTTGTAGAGCAAGTAAAGGCAAGTAAAGGCAAGTAAAG 3949
Db 10060 TTGGGCTGCAAAAGCCCTTGTAGAGCAAGTAAAGGCAAGTAAAGGCAAGTAAAG 10001
QY 3950 CAGATTAATCTTTGATGAACAGGTTAAAAATCAGGCAACATTAAGCATGATGCGGTT 4009
Db 10000 CAGATTAATCTTTGATGAACAGGTTAAAAATCAGGCAACATTAAGCATGATGCGGTT 9941
QY 4010 TTGGCCATTAATCATGGGCGTATTAATTTGGTTAAACGTTGGGGAAGGGAATAGTG 4069
Db 9940 TTGGCCATTAATCATGGGCGTATTAATTTGGTTAAACGTTGGGGAAGGGAATAGTG 9881
QY 4070 CGAGTAAATTAATGCTGAAGAAACAAAGCCGAAATTCATCGAAAAAGCATTAATTAATGCG 4129
Db 9880 CGAGTAAATTAATGCTGAAGAAACAAAGCCGAAATTCATCGAAAAAGCATTAATTAATGCG 9821
QY 4130 TGAATGCAAGTATTCGTTCAATTTAGGCAATTTGGGATTTACAGCTTTATTTGAGATTA 4189
Db 9820 TGAATGCAAGTATTCGTTCAATTTAGGCAATTTGGGATTTACAGCTTTATTTGAGATTA 9761
QY 4190 ATGCGTATTTATGAGAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4249
Db 9760 ATGCGTATTTATGAGAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9701
QY 4250 GCGTTGCAATTAATGCTTAATTAATGCTGAGTACGCGGTGATTAATGCTTTAATCCCGCAG 4309
Db 9700 GCGTTGCAATTAATGCTTAATTAATGCTGAGTACGCGGTGATTAATGCTTTAATCCCGCAG 9641
QY 4310 AGAATATGAGGTTAAGCCTTATTTCTGCAATTTAGTGAATTTGATTTGCAAGCCTTAAG 4369
Db 9640 AGAATATGAGGTTAAGCCTTATTTCTGCAATTTAGTGAATTTGATTTGCAAGCCTTAAG 9581
QY 4370 TACAACCACTGTAATTCGCGGCTGTAACAACCAATTTGAAGCTTAATTTGGCAAAAAG 4429
Db 9580 TACAACCACTGTAATTCGCGGCTGTAACAACCAATTTGAAGCTTAATTTGGCAAAAAG 9521

| Accession | Definition | Accession | Definition |
|-----------------|---|-----------|----------------------------|
| AY 4430 | AAGTGGATTTAAAGCGGAATTTTACATTTCCATTTCCACTTTCGCTTTTATTTCTAAATCTC | 4489 | |
| Db 9520 | AAGTGGATTTAAAGCGGAATTTTACATTTCCATTTCCACTTTCGCTTTTATTTCTAAATCTC | 9461 | |
| QY 4490 | AAGTTCGCAACTCGGTAAACAGCGAATATATGCGGTGAATATAGATATGTTGGTAA | 4549 | |
| Db 9460 | AAGTTCGCAACTCGGTAAACAGCGAATATATGCGGTGAATATAGATATGTTGGTAA | 9401 | |
| QY 4550 | AATCAACTTA-----TTTAT | 4566 | |
| Db 9400 | AATCAACTTAATTTGATCTTTATTTGATTAACAAAGTGGGCGAGATCCACTTTTAT | 9341 | |
| QY 4557 | TCTAATAATGGAACCTTTATTTAATTAATAAGTATCTAATAGACACC--TATAGGGATTA | 4624 | |
| Db 9340 | TCTAATAATGGAACCTTTATTTAATTAATAAGTATCTAATAGACACC--TATAGGGATTA | 9281 | |
| QY 4625 | ATTAGAGATTTAATATGATTTAATTAATTAATTTTCCGCACTTCTGCTGCAGTCG | 4684 | |
| Db 9280 | ATTAGAGATTTAATATGATTTAATTAATTAATTTTCCGCACTTCTGCTGCAGTCG | 9221 | |
| QY 4685 | TATATCTGCTTGTGCAAGAAGTATGACCTGGAATGACAAATCATCTCGCAATATAGCTG | 4744 | |
| Db 9220 | TATATCTGCTTGTGCAAGAAGTATGACCTGGAATGACAAATCATCTCGCAATATAGCTG | 9161 | |
| QY 4745 | AAATGCAACACTTCCACATCTCACTGATTAATAACAGTTGTATTTCTTGCAATTAACAA | 4804 | |
| Db 9160 | AAATGCAACACTTCCACATCTCACTGATTAATAACAGTTGTATTTCTTGCAATTAACAA | 9101 | |
| QY 4805 | CTGTGACTGCACTGATCATTTG 4828 | | |
| Db 9100 | CGGTGACTGCTGTGTATCATTTG 9077 | | |
| RESULT 4 | | | |
| AF517153 | Haemophilus influenzae biotype aegyptius strain HK871 hap gene, complete sequence. | 4054 bp | DNA linear BCT 08-JUL-2002 |
| ACCESSION | AF517153 | | |
| VERSION | AF517153.1 | | |
| KEYWORDS | GI:21666513 | | |
| SOURCE ORGANISM | Haemophilus influenzae biotype aegyptius. Haemophilus influenzae biotype aegyptius Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus. | | |
| REFERENCE | 1 (bases 1 to 4054) | | |
| AUTHORS | Kilian M., Poulsen K. and Lomholt H. | | |
| TITLE | Evolution of the paralogous hap and iga genes in Haemophilus influenzae: evidence for a conserved hap pseudogene associated with microcolony formation in the recently diverged Haemophilus aegyptius and H. influenzae aegyptius | | |
| JOURNAL | Unpublished | | |
| REFERENCE | 2 (bases 1 to 4054) | | |
| AUTHORS | Kilian M., Poulsen K. and Lomholt H. | | |
| TITLE | Direct Submission | | |
| JOURNAL | Submitted (03-JUN-2002) Medical Microbiology and Immunology, University of Aarhus, The Bartholin Building, Aarhus DK-8000, Denmark | | |

Matches 3219: Conservative 0: Mismatches 805 Indels 292: Gaps 14

| | | | |
|----|------|--|------|
| QY | 271 | CGATATTTTAAATACGAAAAATTTACTTAATTAATAACATTATG--AAAAAACTGTAT | 328 |
| Db | 1 | CGATATTTTAAATACGAAAAATTTACTTAATTAATAATAATTTTAAAGAAAAAACTGTAT | 60 |
| QY | 329 | TTCGTGGAATTTTAAACCGCTTCGATTTCTATTTGGAGATAGTATCCGAAGCGGGGACG | 388 |
| Db | 61 | TTTGTGGAATTTTAAACCGCTTTGATTTGATTTGAGATATATTCGAAAGCTTTGGACAG | 120 |
| QY | 389 | GTCAACTATATTTTGGGATGTACTACCAATATATATCGTATTTTCCGAGAAATAAAGGA | 448 |
| Db | 121 | GTCAACTATATTTTGGGATGTACTACCAATATATATCGTATTTTCCGAGAAATAAAGGA | 180 |
| QY | 449 | AGTTCACAGTTGGGGCTAAAAATATTGAGGTTTACAAATAAAATGAAATTTTACTGGCA | 508 |
| Db | 181 | AGTTCACAGTTGGGGCTAAAAATATTGAGGTTTATACAAAAAGGACATTATGTTGGCA | 240 |
| QY | 509 | CATCAATGACAAAGGCCCAATGATGATTTTCCGTGGTCCGCAATGGGGTGGCGG | 568 |
| Db | 241 | CATCAATGACAAAGGCCCAATGATGATTTTCCGTGGTCCGCAATGGGGTGGCGG | 300 |
| QY | 569 | CATTGGTGGGGCATCAGTATATTGAGTGTGGCATAAATGTAGGCTATACCAATGTGG | 628 |
| Db | 301 | CATTAGTAGGGCGATCAGTATATTGAGTGTGGCATAAAGGGGATATACCAATGTGG | 360 |
| QY | 629 | ATTTTGGTGTGAGACAAAAATCCGTGATCAACATCGTTTACTTATATAAATGTGAAAC | 688 |
| Db | 361 | ATTTTGGTGTGAGACGAATCCGTGATCAACACACCGCTTTACTATCAAAATTTAAAAA | 420 |
| QY | 689 | GGAAATATTATA--AAACGATCAAGGCACTCTTATGAGAAAGACTACCACACCCAC | 745 |
| Db | 421 | GGAAATATTATCAGCGTTGGGAGAGAGACATCTTATGAGAGGATATATCATATCCAC | 480 |
| QY | 746 | GCTACATAAATTTGTTAGGAAAGCAOCCCAATGATATAGTTCGTGATATGAAAGGCA | 805 |
| Db | 481 | GTTTACATTAATTTGTAACCGAAGCTGAACCGATAGGTATATA--CTATCTGAGCGTGT | 537 |
| QY | 806 | ACAAATATACAGATAGGACGAAATATCCGAAACGCGTCCGATGCGGCGGTGGCAGT | 865 |
| Db | 538 | GCATATAGTTCAGAGACACCAATATTGGGCTACGGATAAAGATG----- | 582 |
| QY | 866 | TTTGGCGAAACGATCAAAACACGCGGACCAAGTTCGCCGCGCATATCATTTACTGCACG | 925 |
| Db | 583 | ----AGAAACAAATGATATAGTTCATAGGATATTTTCAGAGCGGTATTAACATACCTTAT | 638 |
| QY | 926 | CAGGCAATACACACCAACCAAGCGGAGGAGCGGGCGGTGGTCAAGTCGTGAGCGGCATG | 985 |
| Db | 639 | CAGGAAATACACTACTCAAGTATGTTGGGATATATATGTTACGATTTTATAGCGGTAT | 658 |
| QY | 986 | TGCGCCACAGGGGCAATTACGGGCCCATTCCTATTGACAGGCTCAAGCGGCGACAGCGTT | 1045 |
| Db | 699 | TAAATAGACCTAATCATTTATGTGTCATTTACCTAT--AGGAGCTCAGAGGTATATGTGCT | 755 |
| QY | 1046 | CGCGTATGTTATTTATTTGATGGGAAAAACAAATGGTGTATTAACGGGCTATTATGGA | 1105 |
| Db | 756 | CACCAATGTTATTTATTTGATGGGAGAAACAAATATGGTTTATTAACGGGGTATTATCAAA | 815 |
| QY | 1106 | CCGGCAACCCCTTGGGCGGGGACAGAAATATCATTTCCAACTGTTACGCAAGTCTTTTGG | 1165 |
| Db | 816 | CTGGACATCTCTTTTGGGGAGAGGTAAAGGGTTTCACGTAAATCGTGAAGAAATGGTTT | 875 |
| QY | 1166 | ATGAAATCCCTGAAAAAGATTTGCGTACATCGTTTATATAGCCCATCGGGCAATGGTCAAT | 1225 |
| Db | 876 | ATACAGATTTTAGCAG-----TTGATACCCCTAGTGTTTTTCGACGCT | 920 |
| QY | 1226 | ACACCAATACAGACAAGGCGAGCGGACGCGCATGTCAACACAAACAAAGAGAACCAT | 1285 |
| Db | 921 | ATATTCCTCAATAAATATGACATATATTCCTTTGATCAAAATATATGATGTATACAGTAAAT | 980 |
| QY | 1286 | CTGAAGTCCGCATCGGTTTAAAAAGACAAATTAATCTGCGGAAGGTAAAGACAGATGTTT | 1345 |
| Db | 981 | TAACTTTAACTAACCTAGTAAAGATG-----GCTCTAAAGCTAAATC | 1023 |

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 6, 2003, 00:47:41 ; Search time 930.919 Seconds
(without alignments)
11679.481 Million cell updates/sec

Title: PCT-US03-052226-16

Perfect score: 4828
Sequence: 1 tgcacgcaccttcgagagaa.....gactgcagctgatacatttg 4828

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq.101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|--------|--------------------|---------|-------------|
| 1 | 3039 | 62.9 | 1830121 | HAAT42063 |
| 2 | 2970 | 61.5 | 4319 | HAAT1215 |
| 3 | 1177.6 | 24.4 | 4350 | AAZ1252 |
| 4 | 1108 | 22.9 | 4407 | AAZ1253 |
| 5 | 1083.4 | 22.4 | 4374 | AAZ38918 |
| 6 | 1081.8 | 22.4 | 4374 | AAZ1251 |
| 7 | 1081.8 | 22.4 | 4374 | AAZ1587 |
| 8 | 1081.8 | 22.4 | 4374 | AAA81302 |
| 9 | 1081.8 | 22.4 | 26778 | AAA81477 |

| | | | | | | | |
|---|----|--------|------|--------|----|----------|--------------------|
| C | 10 | 1081.8 | 22.4 | 349980 | 21 | AAZ21612 | Neisseria meningit |
| C | 11 | 1081.8 | 22.4 | 837096 | 21 | AAA81489 | N. meningitidis pa |
| C | 12 | 1041.2 | 21.6 | 4365 | 21 | AAZ38917 | Neisseria meningit |
| C | 13 | 820.8 | 17.0 | 969 | 18 | AAZ88142 | H. influenzae Hap |
| C | 14 | 402.2 | 8.3 | 1332 | 21 | AAA81400 | N. meningitidis Me |
| C | 15 | 401 | 8.3 | 2991 | 20 | AAZ12250 | Neisseria meningit |
| C | 16 | 213.2 | 4.4 | 607 | 17 | AAZ26893 | Haemophilus influe |
| C | 17 | 193 | 4.0 | 5091 | 11 | AAQ06164 | The 19a gene, enco |
| C | 18 | 183.8 | 3.8 | 30078 | 21 | AAQ06164 | N. meningitidis pa |
| C | 19 | 183.8 | 3.8 | 349980 | 21 | AAZ21608 | Neisseria meningit |
| C | 20 | 183.8 | 3.8 | 143768 | 21 | AAA81490 | N. meningitidis B |
| C | 21 | 170 | 3.5 | 781 | 21 | AAA81402 | N. meningitidis Me |
| C | 22 | 169.6 | 3.5 | 891 | 21 | AAA81401 | N. meningitidis Me |
| C | 23 | 155.4 | 3.2 | 4296 | 21 | AAZ54326 | Neisseria meningit |
| C | 24 | 155.4 | 3.2 | 16526 | 21 | AAA81472 | N. meningitidis pa |
| C | 25 | 155.4 | 3.2 | 172325 | 21 | AAZ21613 | Neisseria meningit |
| C | 26 | 155.4 | 3.2 | 349980 | 21 | AAZ21612 | Neisseria meningit |
| C | 27 | 155.4 | 3.2 | 837096 | 21 | AAA81489 | N. meningitidis pa |
| C | 28 | 156.2 | 2.6 | 4113 | 22 | ABA89172 | Escherichia coli p |
| C | 29 | 126.2 | 2.6 | 48254 | 22 | ABA89141 | Escherichia coli p |
| C | 30 | 126.2 | 2.6 | 48345 | 22 | ABA89142 | Escherichia coli p |
| C | 31 | 119.8 | 2.5 | 4128 | 22 | ABA88994 | Escherichia coli p |
| C | 32 | 119.8 | 2.5 | 7654 | 22 | ABA88991 | Escherichia coli p |
| C | 33 | 71.6 | 1.5 | 3399 | 17 | AAQ07587 | Chicken leucocyte |
| C | 34 | 60.2 | 1.2 | 1686 | 16 | AAQ07587 | DNA encoding leuco |
| C | 35 | 60 | 1.2 | 642 | 21 | AAA81729 | N. meningitidis pa |
| C | 36 | 56.6 | 1.2 | 7597 | 24 | ABL33013 | Human immune syste |
| C | 37 | 56.4 | 1.2 | 10595 | 22 | AAZ46542 | Tumour suppressor |
| C | 38 | 55 | 1.1 | 580 | 21 | AAA81832 | N. meningitidis pa |
| C | 39 | 54.6 | 1.1 | 12592 | 24 | AAZ61102 | Human gene regulat |
| C | 40 | 54.2 | 1.1 | 5244 | 24 | ABK40014 | Human chemically p |
| C | 41 | 54 | 1.1 | 3579 | 21 | AAZ70099 | Human chemically p |
| C | 42 | 53 | 1.1 | 10250 | 24 | ABN60083 | Human immune syste |
| C | 43 | 53 | 1.1 | 11422 | 24 | ABK39337 | Human immune syste |
| C | 44 | 53 | 1.1 | 11422 | 24 | ABL32219 | Human immune syste |
| C | 45 | 52 | 1.1 | 7143 | 24 | ABL32983 | Human immune syste |

ALIGNMENTS

RESULT 1
HAAT42063/C
ID HAAT42063 standard; DNA: 1830121 BP.

XX AAAT42063;

DT 14-SEP-1999 (first entry)

DE Haemophilus influenzae complete genome sequence.

KW Genome; bacterium; Haemophilus influenzae; computer readable medium;

KW expression; modulating fragment; regulation; gene expression; vector;

KW organism; open reading frame; ORF; ds.

XX OS Haemophilus influenzae.

XX PN W09633276-A1.

XX PD 24-OCT-1996.

XX PF 22-APR-1996; 96WO-US05320.

XX PR 07-JUN-1995; 95US-0487429.

XX PR 21-APR-1995; 95US-0426787.

XX PR 07-JUN-1995; 95US-0476102.

PA (HUMA-) HUMAN GENOME SCI INC.

XX (UYCO) UNIV JOHNS HOPKINS.

XX PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;

DR WPL: 1996-485782/48.

| | |
|-------------|---------------------------------|
| RESULT 2 | |
| AAT17215 | |
| ID | AAT17215 standard; DNA; 4319 BP |
| XX | |
| AC | AAT17215; |
| XX | |
| 01-JUN-1996 | (first entry) |

| Accession | Gene | Location/Qualifiers |
|-----------|--|---------------------|
| XX | Adhesion and penetration protein gene. | |
| DE | | |
| XX | Adhesion and penetration protein; hap gene; promoter; terminator; | |
| KM | inverted repeat; stem-loop; protease; outer membrane protein; | |
| KM | beta domain; secretion; recombinant vaccine; monoclonal antibody; | |
| KW | diagnostic; immunoassay; ds. | |
| XX | | |
| OS | Haemophilus influenzae. | |
| XX | | |
| EH | Key | Location/Qualifiers |
| FT | -35_signal | 5..10 |
| FT | | /*tag= a |
| FT | -10_signal | 26..31 |
| FT | | /*tag= b |
| FT | sig_peptide | 60..134 |
| FT | | /*tag= c |
| FT | CDS | 60..4244 |
| FT | | /*tag= d |
| FT | mat_peptide | 135..4241 |
| FT | | /*tag= e |
| FT | mat_peptide | 135..2936 |
| FT | | /*tag= f |
| FT | mat_peptide | 2937..4241 |
| FT | | /*tag= g |
| FT | terminator | 4284..4302 |
| FT | | /*tag= h |
| FT | repeat_region | 4284..4302 |
| FT | | /*tag= i |
| FT | stem_loop | 4284..4302 |
| FT | | /*tag= j |
| XX | | |
| XX | W09605858-AL. | |
| XX | | |
| XX | 29-FEB-1996. | |
| XX | | |
| XX | 16-AUG-1995; | 95WO-US10661. |
| XX | | |
| XX | 25-AUG-1994; | 94US-0296791. |
| XX | | |
| PA | (STRD) UNIV LELAND STANFORD JUNIOR. | |
| PA | (UNIW) UNIV WASHINGTON. | |
| XX | | |
| XX | Falkow S, St Geme JW; | |
| XX | | |
| DR | WPI: 1996-151147/15. | |
| DR | P-PSDB; AAR92768. | |
| XX | | |
| PT | Haemophilus adhesion and penetration protein and corresponding DNA | |
| PT | - used to produce vaccines against H. influenzae infection | |
| XX | | |
| PS | Claim 5; Fig 6; 105pp; English. | |
| XX | | |
| CC | The sequence encodes a Haemophilus influenzae adhesion and | |
| CC | penetration protein. The sequence (hap gene) includes putative | |
| CC | -10 and -30 sequences and a putative rho-independent terminator 3' | |
| CC | to the hap stop codon. The terminator contains interrupted | |
| CC | inverted repeats, with the potential for forming a hairpin | |
| CC | structure containing a loop of 3 bases and a stem of 8 bases, | |
| CC | followed by a stretch rich in T residues. The gene product is | |
| CC | first synthesised as a preprotein, which is transported to the | |
| CC | periplasm, followed by insertion of the C-terminal beta-domain into | |
| CC | the outer membrane, possibly forming a pore, and export of the | |
| CC | C-N-terminal fragment through the outer membrane, followed by | |
| CC | autolysis and cleavage and secretion of the mature protease, leaving | |
| CC | an outer membrane protein fragment. The gene may be inserted in a | |
| CC | vector and expressed in recombinant host cells, for use as a | |
| CC | recombinant vaccine. The gene product may also be used in | |

CC diagnostic monoclonal antibody production.

XX Sequence 4319 BP; 1497 A; 776 C; 891 G; 1155 T; 0 other:

Query Match 61.5%; Score 2970; DB 17; Length 4319;

Best Local Similarity 82.3%; Pred. No. 0;

Matches 3581; Conservative 0; Mismatches 645; Indels 124; Gaps 10;

QY 253 TCATAGTCGTTTACCACTATTTTATAGAAAAATACTTATTAATAAACATT 312
 |||||
 Db 1 TCATAGTCGTTTACCACTATTTTATAGAAAAATACTTATTAATAAACATT 59

QY 313 ATGAAAAAACTGATTCGTCGATTTTAAACGCTTGATTCATAGAGATAGTA 372
 |||||
 Db 60 ATGAAAAAACTGATTCGTCGATTTTAAACGCTTGATTCATAGAGATAGTA 119

QY 373 TCCGACACCGTGGCAGCTACTTATTTGGGATGACTACCAATATTCGTGATTT 432
 |||||
 Db 120 TCCGACACCGTGGCAGCTACTTATTTGGGATGACTACCAATATTCGTGATTT 179

QY 433 GCGAGATATAAGGAGATTCACAGTTGGGCTAAAAATATGAGTTTACATAAATA 492
 |||||
 Db 180 GCGAGATATAAGGAGATTCACAGTTGGGCTAAAAATATGAGTTTACATAAATA 239

QY 493 GGAATTTACTTGGCATATCAATGACAAAACCCCAATGATTTTCCGTGCTG 552
 |||||
 Db 240 GGAATTTACTTGGCATATCAATGACAAAACCCCAATGATTTTCCGTGCTG 299

QY 553 CGAATGGGTGGCGCATTTGTGGCGATCAGTATATTTGATGAGTCACATATGTA 612
 |||||
 Db 300 CGAATGGGTGGCGCATTTGTGGCGATCAGTATATTTGATGAGTCACATATGTA 359

QY 613 GGTATACCAATGTGATTTTGTGCTGAGACAAAATCCTGATCAACATCGTTTACT 672
 |||||
 Db 360 GGTATACCAATGTGATTTTGTGCTGAGACAAAATCCTGATCAACATCGTTTACT 419

QY 673 TATAAATGTGAACGGAATTAATTATTAACGATCAACGATCGCTTATGAGAAAG 732
 |||||
 Db 420 TATAAATGTGAACGGAATTAATTATTAACGATCAACGATCGCTTATGAGAAAG 479

QY 733 TACCAACAACCGCTTACATTAATTTGTTCGAGACCAACCCCAATCGATGACTTCT 792
 |||||
 Db 480 TACCAACAACCGCTTACATTAATTTGTTCGAGACCAACCCCAATCGATGACTTCT 539

QY 793 GATATGAACGGCAACAATATACAGATAGGACGAATATCCGACGCGCATGAGG 852
 |||||
 Db 540 GATATGAACGGCAACAATATACAGATAGGACGAATATCCGACGCGCATGAGG 599

QY 853 TCCGAGTGGCATTTTGGCGAAGCATCAAAACGCGACCAAGTTGCCGCGCATAT 912
 |||||
 Db 600 TCCGAGTGGCATTTTGGCGAAGCATCAAAACGCGACCAAGTTGCCGCGCATAT 659

QY 913 CATTAAGTGAAGAGAAATACACACCAACGCGAGAGAGGGCGGTGCTCAGT 972
 |||||
 Db 660 CATTAAGTGAAGAGAAATACACACCAACGCGAGAGAGGGCGGTGCTCAGT 719

QY 973 CTGAGCGCGCATTTGCGCCAGCGGCAATACGCGCCCATTCCTATTCGAGCTCAGC 1032
 |||||
 Db 720 TTGGAGGCGCATTTGCGCCAGCGGCAATATGTCATACGATTCGAGCTCAGC 779

QY 1033 GCGCAGACGCTTGGCTTATTTATGATCGGAAAAACAAAATGCTTATTAAC 1092
 |||||
 Db 780 GCGCAGACGCTTGGCTTATTTATGATCGTAAAAACAAAATGCTTATTAAT 839

QY 1093 GCGCATGAGACCGGCAACCTTGGCGGGGAGACAGATTAATTCATCTGCTGCTG 1152
 |||||
 Db 840 GCGCATGAGACCGGCAACCTTGGAGGCAAGAAATGGCTTCAATGCTGCTG 899

QY 1153 AAGTCTTTTATGATGAATCTTGAATAAGATTTGCTACATCGTTTATAGCCATCG 1212
 |||||
 Db 900 AAGTCTTTTATGATGAATCTTGAATAAGATTTTACATACATCCTTACACCGGACT 959

QY 1213 GGCATATGTCATACCATTAACAGACAAAGCGACGCGCATTTGTCAACACAA 1272

Db 960 GGTATGAGAGTGTACACATTAATAGTGAATATATGTCAGGGGTCTATTAACACGAA 1019
 |||||

QY 1273 ACAGAAACCATCTGAAAGTCCGATCGGTTTAAAGACGAAATTAACCTGCCAGAGT 1332
 |||||

Db 1020 TCAGGAATACCATCAGAAATTAATTAATAGCTAGCAAAATATAGAGTTGCTTGAAG 1079

QY 1333 AAAGAGATGTTTACCA-----ATACAGAGTCCAAATATATACCTGCTGTTGAT 1386
 |||||

Db 1080 AAGGATTAAGTCAATACCTAGATATACGAGCTTAATTTATTTCTCCAGTTTAAAC 1139

QY 1387 AACGTGGAACCTGATTTGAGATCAAAAAACGCGACTGTATCTTCAACCAAC 1446
 |||||

Db 1140 AATGAGAAACGCTATATTTATGATCAAAAAACAGGATCATTAATCTGCAATCTGAC 1199

QY 1447 ATCAACCAAGGTGGCGGCTTGTATTTGAGGTAATTAAGTATAGCTTCTTCAAAAT 1506
 |||||

Db 1200 ATTAACCAAGGTGGCGGCTTGTATTTGAGGTAATTAAGTATAGCTTCTTCAAAAT 1259

QY 1507 AATGCAACTTGGCAAGTGTGAGTGCATGTAGTGAAGACAGTACTGTACTTGAAA 1566
 |||||

Db 1260 AACCAAACTTGGCAAGAGAGCTGCAATACATGTAAGTGAATAAGCACCGTTACTTGAAA 1319

QY 1567 GTAATGTGTGAATAATGATGCTTCTTAATTCGCAAGGACATTCGAGCTTAA 1626
 |||||

Db 1320 GTAATGTGTGAATAATGATGCTTCTTAATTCGCAAGGACATTCGAGCTTAA 1379

QY 1627 GCCAAAGGGGAAATAAAGTTCGATCAGCGTAGCGATGTAAGTCAATTTTGGAGAG 1686
 |||||

Db 1380 GCCAAAGGGGAAATAAAGTTCGATCAGCGTAGCGATGTAAGTCAATTTTGGAGAG 1439

QY 1687 CAGGACAGATTCAGAGCAGCAACAAAGCCTTATGAAATGCTGTTAGTGGCAGA 1746
 |||||

Db 1440 CAGGACAGATTCAGAGCAGCAACAAAGCCTTATGAAATGCTGTTAGTGGCAGA 1499

QY 1747 GGTACGTTGCTGATTAACGATCAACAGCAATTAATACGATTAATTTATTCGGCTTC 1806
 |||||

Db 1500 GGTACGTTGCTGATTAACGATCAACAGCAATTAATACGATTAATTTATTCGGCTTC 1559

QY 1807 CGTGTGTGCTGCTTATGATTTATGAGGCTTCAATTAACCTTTAAACGATTCGAAATACG 1866
 |||||

Db 1560 CGTGTGTGCTGCTTATGATTTATGAGGCTTCAATTAACCTTTAAACGATTCGAAATACG 1619

QY 1867 GATGAGGAGCAACGATTTGATATCAATGACAAAGCAAGATTCACAGTACGATTA 1926
 |||||

Db 1620 GATGAGGAGCAACGATTTGATATCAATGACAAAGCAAGATTCACAGTACGATTA 1679

QY 1927 GGCAGGATACCATTAATACCAACACTGCGGATTAACCAATTAACGATTA----- 1977
 |||||

Db 1680 GGCAGGATACCATTAATACCAACACTGCGGATTAACCAATTAACGATTA----- 1739

QY 1978 ---ATGCTTTTATAGTGTGTTGTTGATTAAGATGATTAATAAATCTGAGAGTTTG 2034
 |||||

Db 1740 GAAATTTCTTACACGCTTGTGTTGCGAAACGATTAATAAATAACATGAGGATTA 1799

QY 2035 AATGTTACTTACATCCGCTTAAACAAGATTAATCACTTCTCTATCAGAGTGAACAAT 2094
 |||||

Db 1800 AATGTTACTTACATCCGCTTAAACAAGATTAATCACTTCTCTATCAGAGTGAACAAT 1859

QY 2095 TTTAAAGGCAATTTACTCAAGAGGTGGCATTGTTAGTGTGAGTGGCCACACCA 2154
 |||||

Db 1860 TTTAAAGGCAATTTACTCAAGAGGTGGCATTGTTAGTGTGAGTGGCCACACCA 1919

QY 2155 CACGATACATCATTTTAA-----TCGCTTAAAGAGCTTGGCGAGCTTAAGGCGAA 2208
 |||||

Db 1920 CACGATACATCATTTTAA-----TCGCTTAAAGAGCTTGGCGAGCTTAAGGCGAA 1979

QY 2209 GTGTTATTTGATGACATGATGATCAACCGTACATTTAAAGCTGAATAATTCGAAATTA 2268
 |||||

Db 1980 ATTTGTTGAGTACATGATGATGATCAACCGTACATTTAAAGCTGAATAATTCGAAATTA 2039

QY 2269 GCGGAGATGCTGCTGTTCTGCAATGTTTCTCAATGAAGAAATGAGCAATTCAGC 2328
 |||||

Wed May 7 15:00:48 2003

pct-us03-05226-16.rng

Page 6

```

Db 2040 GCGGGAATGCGGCTGTTCTCCGAATGTTCTTCAATTGAGGGAATTTGACAGTCAGC 2093
Qy 2329 AATAAGCCAAACGCAATTTGGTGTGGCCAAATTCACAAATATCCATTGGACGGCT 2388
Db 2100 AATATGGAATGCAATTTGGTGTGGCCAAATTCACAAATATCCATTGGACGGCT 2159
Qy 2389 TCAGATTGACAGATTAAAGCACTTGTAAACAGTTAAATTTAACGATAAAAAGTTAT 2448
Db 2160 TCAGATTGACAGATTAAAGCACTTGTCAAAAAGTGGATTAAACGATTAACAAAGTTAT 2219
Qy 2449 GATTCATACCGACACACAAATTAATGCTCTATTAATTTACATATTAAGCAACAGTCG 2508
Db 2220 AATCTATACCAAAACCAATATGCTCTATTAATTTACATATTAAGCAACAGTCG 2279
Qy 2509 AATATTCATGTTTAGCAAAAATTAAATGCTATGCTATTAATTAATCAATGCAATTT 2568
Db 2280 AATGTTAAAGTTAGCAAAAATTAAATGCTATGCTATTAATTAATCAATGCAATTT 2339
Qy 2569 ACATTTAGCAACAAATGCCAACCAACAGCAATATCCAACTTCAATCAACGCAATGCA 2628
Db 2340 ACATTTAGCAACAAATGCCAACCAACAGCAATATGCAATTTGCAATTTCCGCAATTCAG 2399
Qy 2629 AAGGTGATTAAGCAAAATGTAAGGTAAATGCAATTTAAAGGATTTCTGCAATTTCT 2688
Db 2400 AAGGTGATTAAGCAAAATGTAAGGTAAATGCAATTTAAAGGATTTCTGCAATTTCT 2459
Qy 2689 TTAATAAACAGCCATTTTCCACCAATTCAGGGGACAAAGACACAAAGTCAGTTG 2748
Db 2460 TTAATAAACAGCCATTTTCCACCAATTCAGGGGACAAAGACACAAAGTCAGTTG 2519
Qy 2749 GAAATAGGACCTTGACCAATGCTAGCATATGCAATTTGAGAAATTTAAGCTTAATAT 2808
Db 2520 GAAATAGGACCTTGACCAATGCTAGCATATGCAATTTGAGAAATTTAAGCTTAATAT 2579
Qy 2809 AGTACTGTTAGTTAAATTCAGCTTATTCAGCTACCTCAATTAATGCGGACAGTCAGC 2868
Db 2580 AGTACTGTTAGTTAAATTCAGCTTATTCAGCTACCTCAATTAATGCGGACAGTCAGC 2636
Qy 2869 CGTTCTATGAGACGGAACACACGCAACATCGGAGAGACATCGTTTCAACACATTTGCA 2928
Db 2637 CGTTCTATGAGACGGAACACACGCAACATCGGAGAGACATCGTTTCAACACATTTGCA 2696
Qy 2929 GTAATGTAATTTAGGGGGAAGGACATTCATTTCACTTCACTTTATTTGGCTAT 2988
Db 2697 GTAATGTAATTTAGGGGGAAGGACATTCATTTCACTTCACTTTATTTGGCTAT 2756
Qy 2989 AAAAGGATTAATTAATTTATCTATGACGCTGAGGCAATTCACATTTAGCTGTGCG 3048
Db 2757 AAAAGGATTAATTAATTTATCACAATGCGTGAAGGCAATTCACATTTAGCTGTGCG 2816
Qy 3049 GACACAGGCAAAAGACCTGTGACCTTGACCAATTTAACTTTAATGAGGCTGTGATAT 3108
Db 2817 AACACAGGCAAAAGACCTGTGACCAATTTAACTTTGTTGAAGCAAGATATAT 2876
Qy 3109 CAACCCCTGCGAGATTAATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTA 3168
Db 2877 CAACCCCTGCGAGATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 2936
Qy 3169 TGGCGTTATGATTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3228
Db 2937 TTAGCTTATTAATTTAATTTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 2996
Qy 3229 CAGGAATTTGCGCAATGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3288
Db 2997 CAGGAATTTGCGCAATGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3056
Qy 3289 CAAGCTGACTGACTTAAATAACAAATAACTGAGGTAAAGGCGGTCAAAA----- 3342
Db 3057 CAAGCTGACTGACTTAAATAACAAATAAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGT 3116
Qy 3343 ----AGAGCGGCTTTCTCTATACCGGCTGATCAAAAGCAGTTAAACGATTTACAGGC 3399
Db 3116 ----AGAGCGGCTTTCTCTATACCGGCTGATCAAAAGCAGTTAAACGATTTACAGGC 3176

```


Db 138 TTTCGCATCACTACCACTACTATCGCACTTTGCCGAAATTAAGGCAAGTTTCAGT 191

| | | | |
|----|------|---|------|
| QY | 459 | TGGGGCTTAAATAATGAGGTTTACAAATAAAAAGGAAATTTAGTTGGCACTCAATGAC | 518 |
| Db | 198 | CGGGGCCAAGAAATTTGAGGTTTACAAAAAAAGGGAGTTGGTGGCAATCAATGAC | 257 |
| QY | 519 | AAAAGCCCAATGATGATTTTCCGTGGTGTCCGAAATGGGGTGGCGCATTTGGTGG | 578 |
| Db | 258 | AAAAACCCCGGATGATTTTTCGTGGTGTCCGAAAGGGGGGCAATTTGGTGG | 317 |
| QY | 579 | CGATGATATATGATGAGTGGTGCACAAATAAATGAGGCTATACCAATGTTGGATTTGGTGC | 638 |
| Db | 318 | CGATCAATATATTTGTGAGGTGGTGCACATAAGGGGGCTATTAACAAGTTGATTTGGTGC | 377 |
| QY | 639 | TGAAGCAAAAAATCCATGACCAATCCGTTTACTATATAAATTTGSAAMCGSAATAATTA | 698 |
| Db | 378 | GGAAAGAAAGNAATCCGATCCAGCCGCTTTTCTTACCAATTTGTAAAGAAATTA | 437 |
| QY | 699 | TAAAAAGCATCAAAAGCATCTTATGAGAAAGACTACCACACCAGCTTACATTAAT | 758 |
| Db | 438 | TAAAGCTTACATTAACACACCTTTTACAAAGGAGTTATACATATCCCGCTTTGCAATAAT | 497 |
| QY | 759 | TATTACGGAAACCAACCCCAATGATATGACTTTGATATAGAGCAACAATAATACAGA | 818 |
| Db | 498 | TCTCACATATCAAACTCTGAAATGACGAGTACATGATGAGGGGAATTAATCTATTCCGA | 557 |
| QY | 819 | TAGAGCAAAATATCCGCAACGGGTGGTATGAGGCTCCGCGGTGGCACTTTTGGCGAAAGCA | 878 |
| Db | 558 | TAAAGAAATATATCCGAGCTGTGTCCGATCGCTCAGACACACCTATTAGCGTTATAGA | 617 |
| QY | 879 | TCAAAACAACGGGACCAAGTGTCCGGCGCATATCATTTACCTACAGCAGGCAATACACA | 938 |
| Db | 618 | TGATGACAAACACGGCATTTATTCCTACACCGGCGCATGTTTAATTTGGGGCAATACACA | 677 |
| QY | 939 | CACCAAGCGCGACACAGGGGGCGGTGGTCAAGTCTGAGGGGATGTCGGCAAGCGG | 998 |
| Db | 678 | TATGCAAGGTTTGGGAAATATAGGGTANTTATGTTGAGGGGGATGTGGCCATGCCAA | 737 |
| QY | 999 | CATTTACGGCCCCATTCCTATTTCAGAGCTCAAGCGGCGACAGGGGTTGGCTATGTTAT | 1058 |
| Db | 738 | CGACTATGGCCCTATGCGAGTTCAGAGTGGCGACGGGACAGGGGTTGCCCATGTTAT | 797 |
| QY | 1059 | TTATGATCGGGAAGAAACAAAAATGTTGATTAAAGGGGTATGAGGACCGGCACACCTTG | 1118 |
| Db | 798 | TTATGCAAAAACAACATTAATAGCTGCTCAAGCGATTTTCAAAACCGCTTACCCTTA | 857 |
| QY | 1119 | GGGGGGGACAAGATATCTCAATCGGTAGCAGAGCTTTTGTGATGAATACCTTGA | 1178 |
| Db | 858 | TTCCGGGACGGAAACCGTTTCCAGGTGATAGCGAAAGTTGGTTCTACAGATGATTTA | 917 |
| QY | 1179 | AAAA---GATTTGCTTCAATCGTTTATAGCCCATCGGGCATAGTGTCATACACATTAC | 1235 |
| Db | 918 | CAGAGCGATACACATACCGCTTTTAAACCGGCGCATAGCGCAATTTTCTCTTAC | 977 |
| QY | 1236 | AGACAAAGGCGACGGCAGCGCATTTGT-----CAACACAAACAGAGAACCATCTGAA | 1290 |
| Db | 978 | ATCCAACAACAACGATCGGGGTAGCTACAGAAACCAACAAAGGTTTCCATCCAAA | 1037 |
| QY | 1291 | GTTCCGA-----TGGGTTTAAAGACCAAAATTAACCTCCGAAAGGTAAAGACA | 1340 |
| Db | 1038 | GCCTTAAGTACAGACATCGCATGTTTACCAATCTTGATGATGAATGATTAAGAAC | 1097 |
| QY | 1341 | TGTTTACCAATACCAAGGTCCAAATATATACCTGCTCTGTTGATATAGGGTGAACCT | 1400 |
| Db | 1098 | AGTTTACGGGAGGGGGGTATTATAGTACGTCGAAGGTTAAACACGGTGAACCT | 1157 |
| QY | 1401 | GTAATTCGAGATCAAAAAAGGCACTGTTCCTTATCAACAACAATCAACCAAGTGC | 1460 |
| Db | 1158 | TTCTTTTATCGATACGGCAACGGCAAACTCTATTCAACAACAATCAACCAAGGCGC | 1217 |
| QY | 1461 | GGCGGTTTGTATTTTGGGGTAACTTTCAGGTATCTTCAAAATATATGCAACTGGCA | 1520 |
| Db | 1218 | GGCGGTTTGTATTTTGAAGTATATTTACGGTCTCGGCTGTAAGAAACACGAACGTTGGCA | 1277 |
| QY | 1521 | AGGTGCTGATGCTATAGGTGAACAACATGTACTGTTACTTTGGAAAGTAATGCTTTGA | 1580 |